

Weekly Report 5.20 – 5.26

Done

1. 与新加坡同学碰头交流，介绍了实验室一些相关项目、参观了大屏等，在日常生活中有一些指导。
2. 粗略浏览今年 Data Journal Awards 获奖作品，排除一些不符合要求的条件然后并分给大家一起协作鉴赏，后面找时间一起讨论。
3. 小朋友们的入门学习：
 - a. 毅超、璟立、马李彬子各自用 D3 实现了一个最基础的交互可视化，并让他们学习 express 框架和 node 入门。
 - b. 在宇辉的帮助下，我们重新搭建了一个项目的脚手架（模板），并开始将 CCTV 项目的两个视图按照这个脚手架的规范进行迁移，熟悉脚手架的结构与规范；
 - c. 督促彬子每天来实验室；督促陆伊学习可视化；
 - d. 此外，陆伊的教授之前提了一些对于 research proposal 的问题，我与马哥稍作讨论，然后我查阅相关文献、给出了一些参考和大致的回答思路，她将在下周去东京与学姐及导师见面细聊相关问题。
4. Foofong 的 idea 讨论了两次，不太完善（感觉是侧重点有点偏了）。绕来绕去还是跟我最开始和他说的想法没差多少（也即是寒假前聊的），未能把 transition 应用上去。下周我争取帮他想清楚了跟他说明白。
5. 加了一些传媒学院师生，对他们那边应用可视化于数据新闻做了些许了解，简单讨论了一些他们的工作；看到了复旦一个作品，针对热点事件可以直接自动生成（当然也是依据一些固定模板通过爬虫采集得到）诸如关键词搜索次数随事件变化折线图、人群画像、不同领域媒体报道频次直方图，并做成了热点事件舆情数据库的小程序。个人觉得这已经是一套成熟小工具了，信息量丰富且抓住了最重要的几个 facets，可视化形式恰到好处。
6. 一些发现：不少数据新闻可视化作品里用到了 *unit visualization* 的技术（特点是将数据集中每一个数据项使用一个单独的视觉元素进行映射；如 who runs China 里面的一个点代表一个代表，然后从不同角度探索这些数据、以及 Nathan Yau 的一些交互式动画作品）。今年 vis 上，Niklas Elmqvist 提到了这种方法以后我才渐渐意识到这种可视化方法的普遍性。对于数据量不大的、但又要强调个体的，这是一种富有表现力的形式，个人觉得非常适合叙事可视化，也很适合做成动画。Niklas 今年的文章其实已经提出了 Unit vis 的设计空间。

To Do

1. 个人反思，近来时间比较碎片化，没能好好思索想法，也只有大量想法碎片未能整合，后面一段时间我想会调整自己的行为来让自己能更完整的思考、实践自己的 idea。
2. 当前可视化与编程学习进度，我现在安排就是毅超、璟立是第一梯队，学习过程中也有不断调整，同时我也在不断学习；后续同学（主要是陆伊、其他国外回来同学以及大一几位同学）只要按照他们的路径走一遍。之前提到了陆伊等做高维数据探索相关，我最近其实在不断整理发现后面会给他们一起阅读消化。

3. 带领新加坡同学参观新华智云；对他们数据、技术以及 daily routine 有所了解后，我可能会投入一些精力并找天溢多聊聊、来找寻可能的合作点。
4. 之前的思路整理继续向前推进。

论文:

1. JVLC2016 Guiding the exploration of scatterplot data using motif-based interest measures 本文用了一些 CV 技术，定义散点图中的 motif，并应用 tf idf 思想来定义散点图图像特征的 interestingness.
2. CHI2019 Measuring the Separability of Shape, Size and Color in Scatterplots 探讨不同的 channel 不会彼此影响感知。本文研究的是标记的形状、尺寸和颜色会影响多标签散点图中的数据解读。众包实验的结果显示形状会显著的影响颜色和尺寸感知。
3. TVCG2012 Subspace search and visualization to make sense of alternative clusterings in high-dimensional data 本文前面描述的问题非常详细，让我觉得好像把我想到过的都做了；但其实还是借用了 subspace interestingness 的现成方法，采用了最简单的 MDS 投影，利用散点图中点的子空间维度相似性和数据拓扑相似性作为探索的引导。文章里提到的不少相关技术我比较欠缺，诸如 MDS 信息丢失我该怎么补充、拓扑空间相似我怎么度量，这些也同样值得我学习。

规划:

短期: data2video 的想法整合、与福风的讨论。月底前初步完成。

短期: 新加坡以及其他小朋友的可视化学习以及编程学习的推进；多接触实际的可视化作品与现有系统代码（未来两周）。

中期: 工程能力方面: 近日集中在 1) D3 方面一些入门的数据绑定范式的学习 2) 算法分析初步学习。

工作时间

平时 50, 周末一共 5 小时. 总共 55 小时.